ORAU

ORAU COMMENTS/RECOMMENDATIONS
PRELIMINARY RESULTS AND RECOMMENDATIONS OF THE INDEPENDENT ASSESSMENT OF PARCEL EVALUATION FINDINGS REPORT FORMER HUNTERS POINT NAVAL SHIPYARD, SAN FRANCISCO, CALIFORNIA

GENERAL DISCUSSION

Per Research Subcontract No. V0784A-A between Oak Ridge Associated Universities (ORAU) and Oregon State University, ORAU performed the initial, high-level review and independent assessment of:

1. *Draft Radiological Data Evaluation Findings Report for Parcels B and G Soil Former Hunters Point Naval Shipyard San Francisco, California*
2. Appendix A: K-S Test Results
3. Appendix C: Parcel B and Parcel G Evaluation Forms

Since the time of ORAU's initial review, the independent verification support contract mechanism has been migrated to Agreement Number 101000961 between ORAU and CH2M.

A limited review was also performed of Parcel B soil sample analytical data available in the Final Radiological Evaluation Database (FRED). The objective of the limited review was to gain familiarity with the analytical data that project evaluators had used as inputs for statistical assessments and their conclusions regarding the potential for survey unit data falsification, and also for cross referencing the laboratory data for several specific samples that were called out in Appendix C Evaluation Forms. In addition, manual searches were conducted of the folders and individual files contained within the shared directory at https://delivery.ch2m.com/projects/684353, again to gain familiarity with available records and data.

ORAU must emphasize that these searches identified what must be described as a massive number of records and page counts of individual files—Appendix C alone contained over 3,000 pages and the FRED query for Parcel B Trench soil sample data contained in excess of 100,000 individual results. The ORAU schedule for review began on September 27 and with a requested completion date of October 13, 2017. This compressed schedule did not allow for preparation of the review's data quality objectives, nor did it allow for completing a thorough independent assessment of the data integrity itself and the methods and decision processes employed to develop an evidentiary conclusion of data manipulation or falsification (as described in the Draft Radiological Data Evaluation Findings Report). Therefore, ORAU offers two primary general observations together with the respective conclusions, and recommendations regarding the Hunters Point data evaluation for Parcels B and G soils. In some cases, the discussions include more specific observations. These are provided as examples for the observations and are not a complete representation of other specific comments identified during the review.

OBSERVATIONS AND CONCLUSIONS

1. EVALUATION METHODS

1.1 Observations

In general, the primary method used of comparing and statistically evaluating the concentrations of ubiquitous naturally occurring radionuclides Ac-228, Bi-214, and K-40 does provide a technically appropriate screening methodology to identify suspect sample populations for which additional assessment is warranted. It was notable that in some (but not all) evaluation narratives, further data exploration occurred, in particular examining the concentrations of other decay chain-related gamma emitters in the case of Ac-228 and Bi-214. Although several of the evaluators did examine another Ra-226 progeny—specifically Pb-214—when Bi-214 results were found, ORAU is uncertain of the basis for selecting Bi-214 as a marker rather than Pb-214 to eliminate the variable of coincidence losses seen with Bi-214 to assess data for indications of falsification/sample replacement. Additionally, the comparisons of the U-238 daughter, Th-234, with the onsite lab's reported Ra-226 concentration may have also provided valuable information on the presence of excess radium in samples resulting from contamination versus simply background variability. It is, of course, possible that the evaluation team considered this approach when developing the methods and determined it was not beneficial to the study goal.

Also, the population assessment methods for comparing intra-survey unit results, i.e., characterization populations with final status survey (FSS) samples, which identified disagreement when comparing FSS data with co-located units, provided additional assurance that the data should be considered suspect. However, further investigation would have been beneficial, although likely not achievable. This further investigation would have involved evaluation as to whether different soil types had been encountered after remediation, in either a portion of or all of a survey unit, from the soil type present during characterization. There was some limited macro discussion of soil type variability among co-located survey units or site wide geology. However, the Appendix C narratives do not discuss that any micro evaluations of the variability of soil types within the survey unit were performed, particularly vertical variability for excavated versus non-excavated survey unit sections.

The statistical tests applied were appropriate. However, there were a number of instances where the methods could be misinterpreted without necessary caveats provided. In several cases, statistical assessments were made on as few as 2 samples. Box plots comparison of means and variabilities and the associated statistical tests appeared to test a population of only 2 to 3 samples against a population of several hundred.

Lastly, the recommendations for no further action, archived sample analysis, or confirmatory sampling and analysis were reviewed, which identified a report deficiency. That is, the report did not provide specific information as to the requirements for demonstrating compliance with the Table 2-1 release criteria. As to the suitability of any of the data for site status decisions or for planning future investigations, compliance information would assist in prioritizing investigations should stakeholder further action recommendations be implemented. More specifically:

- Was compliance based on not-to-exceed thresholds?
- Is a statistical assessment of population means/medians required, such as the Wilcoxon Rank Sum or Sign tests?
- Are there area factors for hot spot assessments?
- Is the unity rule applicable for multiple contaminants?
- Are there other criteria?

The determination of compliance with clean-up objectives or even the data usability remains open to interpretation without specific, quantifiable and objective criteria.

ORAU agrees in most of the cases with the recommendations. However, several survey units were recommended for no further action that ORAU believes should be candidates for further investigation.

1.2 Conclusions

Overall the evaluation methods and, more importantly, the recommendations for further action provide a good faith effort to identify and document candidate survey units or entire site areas that should receive further quantitative evaluations to assess the current radiological status of selected survey units.

1.3 Recommendations

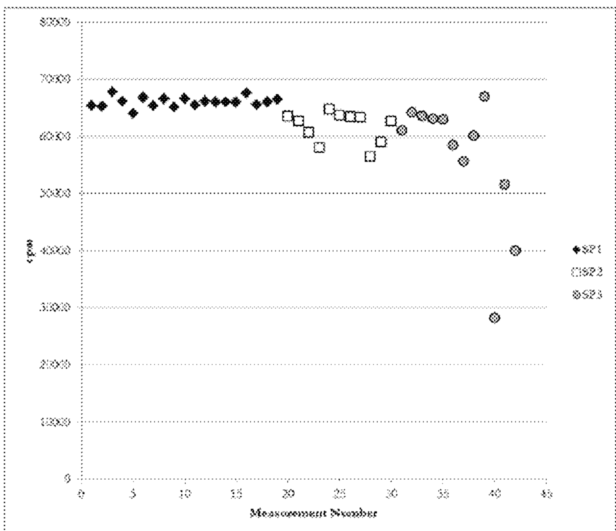
ORAU would recommend additional statistical data population evaluations of a subset of the units that were recommended for further confirmatory sampling. The objective of such an evaluation would be to clearly establish whether discrepancies noted are indeed the result of data falsification/manipulation or, in some cases, the result of different soil types, variability/bias in the onsite laboratory results, or a combination. Section 2 provides further discussion regarding analytical results.

2. Survey And Analytical Performance

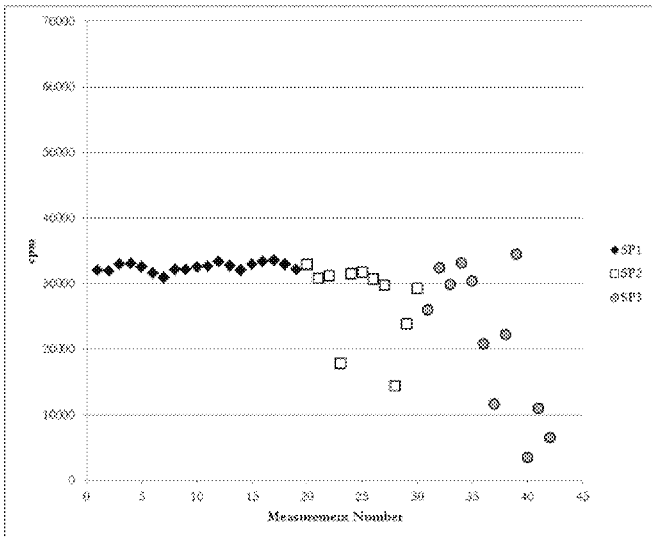
ORAU has gained extensive experience in common procedural and equipment deficiencies while conducting independent verification surveys at hundreds of sites. The review of the evaluation forms and initial reviews of supplementary data, when available, identified symptoms of several commonly observed issues with the remediation contractors' field survey and onsite laboratory processes. If these issues are confirmed, not only are the data suspect for those units identified as potentially manipulated or falsified, then all data are questionable and therefore inadequate for site status decisions. ORAU provides the following supporting observations to substantiate this statement.

2.1 Survey Performance

ORAU was informed that most radiological gamma radiation scan data were not electronically captured; rather, information was documented by recording scan ranges. The scan ranges were provided within the Appendix C evaluation forms when available and germane to a particular survey unit evaluation. The forms did not refer to a specific gamma detector used, rather, references only the model number of a ratemeter-scaler and with no information regarding the detector model or serial number. Based on the background ranges and action levels quoted in the forms, it appears that the general detector response ranged from 5,000 to 9,000 counts per minute (cpm). These background responses are characteristic of a 2-inch by 2-inch NaI(Tl) scintillation detector. ORAU uses similar detectors and has extensive experience with the detector and expected responses. Multiple instances were seen in Appendix C where detector response was as low as 2,000 to 3,000 cpm, while the more common low end of quoted response ranges were 4,000 to 5,000 cpm (which although on the lower tail of the typical response distribution is not unusual). However, the 2,000 cpm values are highly unusual unless: (1) the detector is being operated in a shielded or windowed configuration; (2) the detector being used in those survey units had a crystal smaller than 2×2; (3) the area surveyed was anomalous as exhibiting a very low background; or (4) the detector crystal had degraded and the operating voltage was not set properly. Based on the various count rates reported in Appendix C, the most probable reason is a degraded or otherwise unstable measurement system. As such, there is a possibility that the detector was significantly under-responding, especially to low energy photons. The following figures showing data from an ORAU study on detector performance illustrate this point. The figures represent gamma energy response from three NaI 2×2 populations to mid- and low-range energy photons (Cs-137 and Am-241, respectively). The populations were new detectors (SP1), 2 to 5-year old detectors (SP2), and detectors more than 5-years old (SP3). All were operated at the manufacturer's recommended high voltage. Response degradation is clearly evident.



cpm for Cs-137 by Subpopulation at 900 Volts



cpm for Am-241 by Subpopulation at 900 Volts

If the measurement system was degraded or otherwise unstable, then the mean plus 3 sigma action levels provided in supplementary data packages are likely to be deficient for these detectors. Most action levels were quoted in the 6,000 to 9,000 cpm range. Therefore, the potential for undetected contamination is highest in those survey units where these detectors were used. Additional evaluations would be necessary to confirm the type of detector and stability of the measurement system and affected survey units to make this determination.

2.2 On-Site Laboratory

While reviewing the report and Appendix C and the dynamics of the population testing, evidence was found that that the onsite gamma spectroscopy system was unstable or otherwise not calibrated appropriately. Instability would result in systematic bias during different analytical events and could be another reason why some of the time plots showed differences in the populations. The primary evidence of this is that results for one of the trench survey unit sample populations are almost all negative for Cs-137, while the other sample populations from the same unit show expected results both above and below zero in the plot. Systematic errors in reported concentrations could have resulted from numerous parameters. These include:

- Incorrect efficiency calibration.
- Incorrect energy calibration.
- Counting samples in a non-calibrated geometry.
- Counting samples wet and not correcting for the percent moisture. Concentrations are typically reported on a per dry weight basis. Discussions with evaluation staff indicated that samples were likely not dried prior to analysis.
- Poor sample preparation. Samples should be dried, homogenized, and sieved before counting in a calibrated geometry.
- Analysts not following a sound, approved standard operating procedure (SOP) or having the SOPs change often.
- Incorrect use of the system's software. Examples include (but are not limited to) Peak Tolerance, Identification Energy Tolerance, and ROI Limits Determination.

- Using the incorrect energy line to identify a radionuclide. An example is using the 186 keV peak for Ra-226 when there is U-235 present in the sample, or using the Pb-214 peaks without sufficient time for the radon progeny to grow in. Also, typically the 351 keV peak of Pb-214 is a more reliable peak to use in estimating Ra-226 concentration than the 609 keV peak of Bi-214.

- Samplers or lab analysts switching samples.

- Not subtracting out a correct (and current) environmental background from the analytes counts.

- Faulty detectors, which may lead to drift and poor resolution, among other issues.

- Using the incorrect geometry.

- Not performing daily QC to ensure the detector is sound.

- Not counting the sample in the same environment as the calibration standard was counted.

Deficiencies in even one of the above parameters could introduce significant bias. The simplest example is if samples were not dried, percent moisture would not only have an impact on reported concentration values but when combined with other populations with higher or lesser moisture content, could be erroneously misinterpreted as coming from more than one population. In addition to ORAU independently concluding that there are significant quality assurance issues, several of the Appendix C evaluation forms similarly noted that discrepancies in results were likely the result of quality assurance issues and recommended no further action. Several of these affected units are related to the Observation 1.1 section, where ORAU did not support the no further action conclusion.

If these suspected analytical issues are substantiated, it is ORAU's opinion that most of the reported results, including those samples collected from survey units where no evidence of data falsification was identified, should be considered suspect and potentially unusable.

2.3 Conclusions

ORAU identified potential issues with both field and laboratory data quality during the review. In both instances, the suspected biases observed would negatively impact the data integrity and therefore the data use for radiological status decisions.

2.4 Recommendations

For the survey performance observation in Section 2.1, ORAU recommends a thorough review of the complete instrumentation operational packages for both suspect and non-suspect detector systems. Records for review include electronic calibration, high voltage plateaus, daily operational background and check source responses, and field records representing various use periods for the selected detectors.

The on-site laboratory observation in Section 2.2 would require extensive review of records that may or may not be available. As an initial assessment, ORAU recommends that a representative population of archived samples from various project time periods be selected for independent analysis and comparison of the data with the results generated from the onsite laboratory. ORAU operates the Radiological and Environmental Analytical Laboratory (REAL), an ISO-17025 registered laboratory operation. The REAL performs all confirmatory analysis on behalf of the U.S. Nuclear Regulatory Commission and continuously demonstrates the reliability of reported analytical through multiple performance evaluation programs in which they participate. ORAU radiochemistry experts are available to provide this evaluation to further assess the adequacy of the onsite laboratory data.

INITIAL SUMMARY RECOMMENDATIONS

ORAU recommends a graded approach consisting of two phases be applied to further assess the current site status. Phase 1 would be an assessment of archived samples and Phase 2 a confirmatory survey investigation of both individual and combined suspect and non-suspect units.

Initially, an assessment of representative populations of archived samples is recommended to address a principal study question of whether any site data that Tetra Tech EC, Inc. generated is usable for localized site status decisions. Should stakeholders elect to pursue an archived sample study, the recommendation for the assessment is for the review team and stakeholders to select representative populations of characterization, remedial action support, and final status survey archived soil samples from the most highly suspect trench, fill, and building site survey units. They should also select corresponding populations of samples from survey units for which there was no evidence of potential data manipulation or falsification. The data between the on-site laboratory, off-site laboratory (for any samples in the selected population that were sent off-site), and independent confirmatory analysis laboratory would be evaluated for both individual radionuclide paired measurement results and for the paired unit populations via appropriate statistical tests. This evaluation would serve as an input to the decision as to the usability of analytical data from samples collected in site areas with no or low potential for contamination and where there would otherwise be no motive for falsification/manipulation.

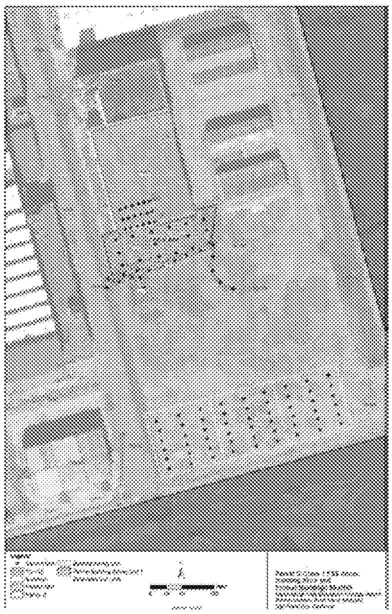
Secondly, develop a robust site investigation and confirmatory sampling and analysis plan designed with input from various stakeholders. The investigation may require multiple investigation events, based on the results obtained from the survey units selected for investigation.

ATTACHMENT:
PROPOSED SOIL APPROACH RECAP AND ORAU COMMENTS/RECOMMENDATIONS

The regulatory agencies proposed to the Navy that to confirm if there is low potential for contamination in the storm/sewer system, then we have to do a "prove-out". They propose re-excavating 25% of the trenches and conducting class 1 surveys of the soil and sidewalls using the past work plans. Presumably this would mean continuing with a point-by-point comparison to the release criteria.

The Navy agrees that reliable data needs to be collected to confirm the Navy's conceptual site model. However, we believe the agencies' proposal is excessive. We have developed an alternative proposal that we think is technically defensible, and will give the parties assurance that site conditions are safe and our remedial action objectives have been met. This alternative proposal has three parts:

1. *Conduct Class 1 surveys of soil and associated trenches where contamination was known to have existed according to the Historical Radiological Assessment. In Parcel G, the only area with known contamination to have existed is associated with former building 364. A pictorial of the approach is shown in FigureX_Parcel_G_Class1_SamplingAreas.pdf, attached.*



Comments/Recommendations:

A. Documentation provided to ORAU, including the various parcel radiological data evaluation findings reports and the *Basewide Radiological Removal Action Memorandum (*BRRAM*)*, does not imply or otherwise state whether the DCGL/risk/dose compliance criteria include application of the unity rule. Rather, the text provides general statements, such as the following in Sections V.A., V.A.1, and V.A.5 of the BRRAM:

   a) "MARSSIM guidance is being used to apply the cleanup goals."

   b) "Soil confirmation sampling and analyses; comparison of results against cleanup goals listed in Table 1 using MARSSIM methodology"

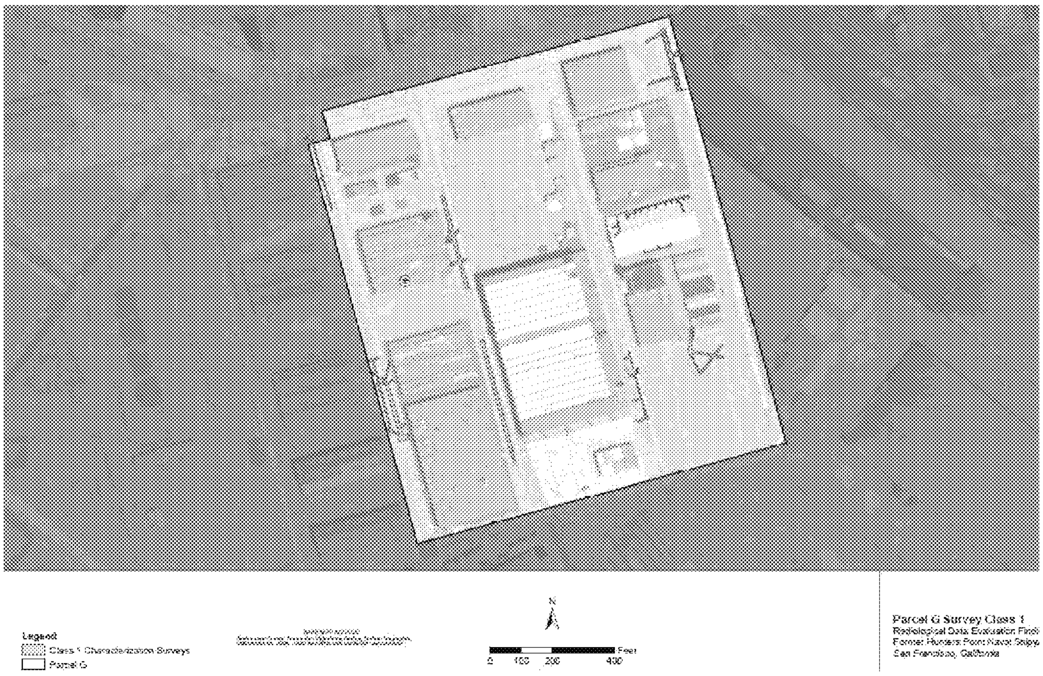
   c) "Application of soils, debris, and surface cleanup goals to sites: MARSSIM guidance"

      i. Per Appendix A of the BRRAM, "The Navy would apply this requirement…"—in reference to the Radiological Criteria for Unrestricted Use at Closing NRC Licensed Facilities—"…with the exception that 15 mrem/yr TEDE is substituted for 25 mrem, as ALARA." As the Table 1 residual doses range from 0.05263 to 25 mrem/yr for the respective release criteria, ORAU assumed a the time of this review that the unity rule could be required to demonstrate compliance. Since that time and participation in the January 17 and 18 technical team, it is ORAU's understanding that the criteria apply independently. Whether the release criteria apply independently, but more importantly if the unity rule is required, the Navy should consider increasing *m* (background reference area) and *n* (survey unit) sample populations from 18 to minimize Type II error probabilities. Otherwise, consider assessing available sample data to estimate the mean and standard deviation (in terms of unity if the unity rule will be applied, or if not, for each primary radionuclides of concern (ROCs) to estimate the worst case relative shift. Next, evaluate the relative shift and corresponding sample size with the proposed *n* and *m* of 18. Prepare a prospective power curve to assess probability of Type II error.

      ii. Additionally, the Class 3 plan specifies a 99% confidence that the concentrations of ROCs do not pose a risk. The plan should specify the confidence level (and associated Type I error) for the Class 1 investigation. Will the planned confidence be the default of 95% (Type I decision error limit of 0.05) or is the intent to also impose 99% confidence (Type I decision error limit of 0.01) in the release decision for Class 1 areas? If the latter, the selected Type I error may also impact the necessary sample size. For example, the proposed sample populations of 18, assuming the default regulator acceptable error 0.05 (95% confidence level) and Type II error of 0.05 corresponds to a relative shift of 1.5. However, if the Type I error is changed to 0.01 (99% confidence) and all other variables are held constant, the population for *n* and *m* increases to 25 to achieve the desired confidence level.

B. Although the rectangular sample pattern is acceptable for this type of planned investigation, the recommendation is that a triangular pattern be substituted to increase probability of identifying residual areas of elevated radionuclide of concern (ROC) concentrations, in particular for survey units where scanning may provide limited information due to backfilling or otherwise assessing subsurface soils.

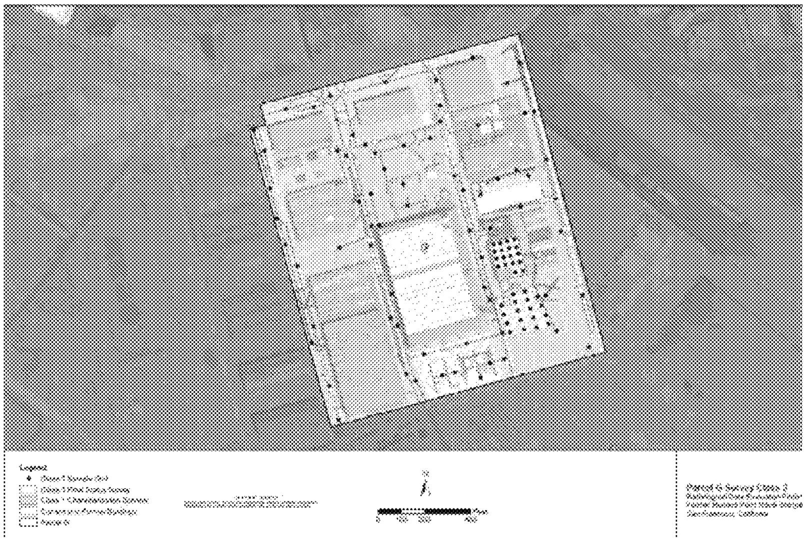
2. *Conduct Class 1 Surface surveys of a representative portion of the soil in question. The purpose of this is to be able to complete 100% scans and sampling and biased sampling of a representative portion of the soil without costly, timely, and unnecessary excavation. Twelve of the 63 trench units in Parcel G were backfilled with only native fill (i.e. all other trenches have some or all import fill and are therefore not representative). These 12 trenches were selected for the Class 1 Surface surveys and are shown on FigureX_Parcel_G_Survey_Class_1.pdf, attached.*



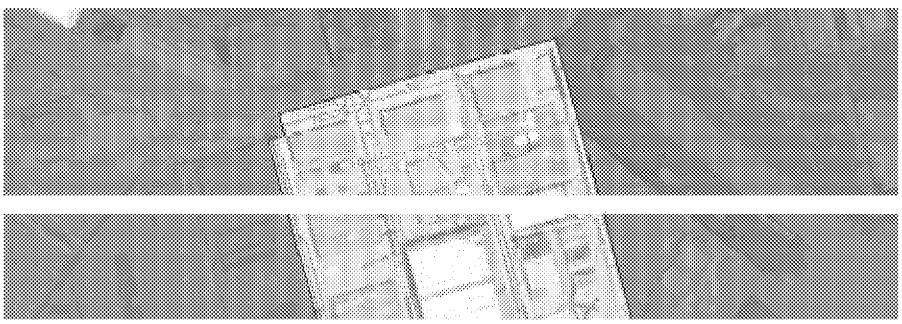
<u>Comments/Recommendations</u>:

A. Refer to comment/recommendations A.i, A.ii, and B above for Proposed Approach Item 1.

3. *Conduct Class 3 Surveys on all of the land areas and trench units. The land areas and trench units with a low potential for contamination would be combined and divided into Class 3 survey units. Eighteen systematic (3 dimensional) sampling points would be distributed throughout the Class 3 survey units. The purpose is to provide a confidence level (~99%) that the concentrations of radionuclides of concern do not pose a risk. FigureX_Parcel_G_Survey_Class_3.pdf shows a preliminary sampling distribution.*



A figure summarizing the overall approach for Parcel G (i.e. combining components 1,2, and 3) is also attached (FigureX_Parcel_G_Survey_Class_Summary.pdf). **(Note: Original figure file corrupted)**



Comments/Recommendations:

A. Refer to applicable components of comment/recommendation A.i above for Proposed Approach Item 1.

Data Analysis:

We believe that it is necessary to update the process to evaluate soil sample results. This means using MARSSIM-based statistical analysis to confirm if the concentrations are acceptable, as opposed to using the point-by-point comparison practices in the past. We would therefore collect a new, site-wide background; use the WRS test to analyze each survey units' systematic results; compare each individual sample result to an EMC (under development but possibly calculated using 10^-4 risk); and completing a NORM evaluation for samples with exceedances. We would also do a final dose and risk calculation of the survey results for each survey unit.

Comments/Recommendations:

Concur application of the statistical testing to demonstrating compliance is a critical step vs. the previous sample-by-sample, not-to-exceed threshold evaluation. Many of the evaluation forms indicated a single sample may have exceeded the DCGL for an individual ROC, generally Ra-226 or Cs-137, leading to significant over-remediation of a survey unit that would have likely readily satisfied a statistical assessment demonstrating that the mean/median concentration was less than the DCGL.

The data analysis discussion does not specify whether the unity rule (DCGL = 1 and sum-of-the-ratios calculated for each analytical result) will be applied prior to performing the WRS test, or if each ROC will be tested independently. The planned approach should be specified.

Consider increasing *m* and *n* (background and survey unit sample populations) to minimize Type II decision errors, especially if applying unity rule.